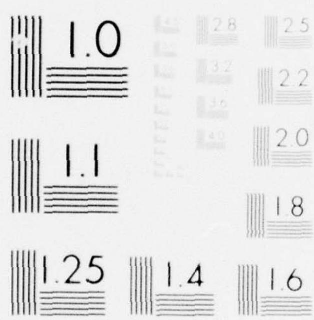
AD-A044 716 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/2
MATHEMATICAL MODELS OF ROAD TRAVEL DISTANCES.(U)
FEB 77 R F LOVE, J G MORRIS DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1719 NL

| OF |

AD
A044716

END
DATE
FILMED
10-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MRC Technical Summary Report # 1719

MATHEMATICAL MODELS OF ROAD TRAVEL DISTANCES

Robert F. Love and James G. Morris

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

See 1473

February 1977

(Received March 19, 1976)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

MATHEMATICAL MODELS OF ROAD TRAVEL DISTANCES

Robert F. Love[†] and James G. Morris[‡]

Technical Summary Report # 1719
February 1977

ABSTRACT

Management science models often require estimates of distances between points in a road network based on the point coordinates. These estimates are typically derived using rectangular or Euclidean distance functions. The present investigation employs more general functions to estimate samples of urban and rural road distances. Proofs of conditions for the convexity of the functions are given. Statistical comparisons are made of prediction accuracy in order to establish relative merit. Sensitivity of the best parameters of the functions with respect to different geographical areas suggests the importance of using empirical distance functions.

AMS (MOS) Subject Classification - 90B99

Key Words - Estimates, Distances, Road, Network, Point, Coordinates, Functions, Urban, Rural, Statistical, Parameter

Work Unit Number 5 (Mathematical Programming and Operations Research)

[†]Research sponsored by the United States Army under Contract No. DAAG29-75-C-0024 and a grant from the Graduate School of the University of Wisconsin. Visiting Professor, University of Waterloo, Waterloo, Ontario, in 1976-1977.

[‡]Research supported by a summer research grant awarded by the Graduate School, Kent State University. Visiting Associate Professor, University of Wisconsin-Madison, in 1976-1977.

MATHEMATICAL MODELS OF ROAD TRAVEL DISTANCES

Robert F. Love† and James G. Morris‡

1. Introduction

The research described here was carried out to fit mathematical functions to sets of urban and rural road distance data. The objective of the study was to evaluate distance predicting functions. In a previous study [6], seven different functions were investigated which transform coordinates of points into estimates of road distances between those points. The function which gave superior accuracy was

$$d(q, r; k, p, s) = k\left[\sum_{i=1}^{2} |q_i - r_i|^p\right]^{1/s}$$

where $q=(q_1, q_2)$ and $r=(r_1, r_2)$ are points in two-dimensional space and k, p and s are parameters. The most suitable value for s was not markedly different from that of p when the function was "fitted" to two different sets

---

†Research sponsored by the United States Army under Contract No. DAAG29-75-C-0024 and a grant from the Graduate School of the University of Wisconsin. Visiting Professor, University of Waterloo, Waterloo, Ontario, in 1976-1977.

‡Research supported by a summer research grant awarded by the Graduate School, Kent State University. Visiting Associate Professor, University of Wisconsin-Madison, in 1976-1977.

of sampled inter-city road distances. The function $d(q, r; k, p, p)$ was suggested for practical use since one less parameter must be estimated from the data.

As many locational analyses are accomplished within an urban or rural setting, it seems appropriate to investigate whether such distances can also be modeled more accurately than with the customary rectangular or Euclidean distance functions. The data sets used in this paper were samples of road distances in a Wisconsin rural region and in the Wisconsin cities of Madison and Milwaukee. In addition, sampled road distances were included from the cities of Canton, Columbus, and Toledo in Ohio. These cities were chosen to provide a spectrum of road network regularity and varying degrees of accommodations to physical obstructions.

Five distance functions are established as models for consideration. These are denoted by:

$$d_1(q, r) = k \sum_{i=1}^{2} |q_i(\theta) - r_i(\theta)|, \; k \geq 0, \text{ where } q_1(\theta) = q_1 \cos\theta + q_2 \sin\theta,$$

$$q_2(\theta) = -q_1 \sin\theta + q_2 \cos\theta \text{ and similarly for } r_i(\theta),$$

$$d_2(q, r) = d(q, r; k, 2, 2), \; k \geq 0,$$

$$d_3(q, r) = d(q, r; k, p, p), \; k \geq 0, \;\; p \geq 1,$$

$$d_4(q, r) = d(q, r; k, p, s), \; k \geq 0, \;\; p \geq 1, \; p \geq s,$$

$$d_5(q, r) = [(q-r)'M(q-r)]^{\frac{1}{2}}, \text{ where } M = \begin{bmatrix} m_1 & m_2/2 \\ m_2/2 & m_3 \end{bmatrix} \text{ is a positive definite}$$

symmetric matrix, $(q-r)$ is taken to be a 2×1 column vector, and a prime denotes the transpose operation.

d_1 is the rectangular distance function with axes rotated through an angle θ from the original coordinate axes, multiplied by the constant k. d_2 is the Euclidean metric. d_3 and d_4 replace the square root and square of the Euclidean metric by 1 and 2 parameters, respectively. d_5 provides for weighting the squared coordinate differences differently as well as a weighted cross-product term within the square root. Unit "circles" given by the points satisfying $d_f(q,0) = 1$, $f=1,2,3$, for $k=1$, $p \geq 1$ are symmetric with respect to orthogonal coordinate axes. However, $d_5(q,0) = 1$ generates a locus of points q at unit distance from the origin which forms an ellipse. This is true since $d_5(q,0) = [q'Mq]^{\frac{1}{2}} = 1$ holds only if $q'Mq = 1$. The locus of points q such that $q'Mq = 1$ forms an ellipse when $m_2^2 < 4m_1m_3$ and this inequality holds since M is positive definite. Consequently, a different form of directional bias in distances between points is modeled using d_5 versus using d_1, say. Specifically, the directional bias inherent in d_5 is such that the travel direction of greatest ease is perpendicular to that of greatest difficulty. For d_1 these directions are at 45° angles to one another. d_5 is invariant under rotation of the axes since the best direction of axes is chosen through the parameter m_2 in accordance with a best-fit criterion. Except for the Euclidean distance function the remaining functions are not invariant under rotation of axes; the analyst's choice of axes affects the ability of d_1, d_3 and d_4 to accurately predict distances. The inclusion of θ in $d_1(q,r)$ allows for a properly oriented coordinate system to study the ability of the rectangular distance function to predict urban road distances. $d_5(q,r)$ has but three parameters which must be estimated since M is symmetric and since any multiplicative constant k can be assumed to have been factored into the values of m_1, m_2, and m_3. The indicated conditions on the parameters are required for convexity. The authors have shown that $d_3(q,r)$ is convex [3].

The results in the appendix establish that $d_4(q,r)$ and $d_5(q,r)$ are also convex.

Empirical distance functions such as d_3, d_4 and d_5 can be used in a wide range of applications. Four specific applications which have come to the authors' attention are discussed here.

1. Vehicle Travel Times.

Kolesar, Walker, and Hausner [5] have formulated and validated a model for predicting fire engine travel time as follows:

$$T(D) = \begin{cases} 2(D/a)^{\frac{1}{2}} & , \text{ if } D \leq 2d_c, \\ V_c/a + D/V_c, & \text{if } D > 2d_c, \end{cases}$$

where T is average fire engine travel time, D is the length of the run, a = acceleration, d_c = distance required to achieve cruising velocity, and V_c = cruising velocity.

It would seem logical that the model or a variation of it could be used for many other types of vehicle travel times; we note that the independent variable of the model is the travel distance. Using the empirical methods we suggest here, given the coordinates of any point in the area to be served and the coordinates of the service center, the travel distance may be generated within the model. When large numbers of points are to be considered or the points are not known in advance (i.e., may be generated randomly or iteratively by the model), the advantages of using an empirical distance function are obvious.

2. Characterization of Road Networks.

The best-fit parameters of the distance functions provide for reduction of large amounts of distance data to a form which makes possible certain generalizations about a road network. Consider $d(q, r; k, p, p)$. If k and p = 1, the network is basically a rectangular grid. As k increases while

$p = 1$ there remains a rectangular bias such that travel is most efficient parallel to the orthogonal coordinate axes and least efficient at 45° angles to the axes; but there is an increasing nonlinearity in the roadway. A continuum of rectangular bias is measured along $1 \leq p < 2$. If $p = 2$, the network is highly developed since efficiency of travel is not significantly affected by directionality. If $p > 2$ the directional bias is inverted since travel is most efficient at 45° angles to the axes and least efficient parallel to the axes.

The implicit assumption here is that the effects of k and p are separable which may not be strictly valid; e.g., k may not be interpretable as the sole measure of "curvature" separate and apart from p. Similar analyses are possible for the remaining distance functions considered in this study.

3. Verifying Distances in Road Networks.

As Ginsburgh and Hansen [3] suggest, it is not uncommon that road network data are unreliable. One of their suggested remedies is to verify the arc length between nodes i and j whenever this length does not lie in the "confidence" interval $[(1-c)d_{ij}^{*}, (1+c)d_{ij}^{*}]$. Here d_{ij}^{*} is the estimated distance and c is heuristically determined. A transportation manager or urban planner might screen road network data using this approach replacing d_{ij}^{*} by one of the estimating functions considered in this study (if node coordinates can be obtained). We have found this to be a useful technique for verifying data for traveling salesman and truck dispatching models.

4. Facilities Location Models.

In facilities location models the objective is to locate private or public units such as factories, distribution centers, ambulance bases, postal stations, etc., to optimize an effectiveness criterion involving distances. Empirical distance functions can improve estimation of road travel distances to hypothetical facility locations which may be anywhere in the two-dimensional continuum or obviate the need to individually measure myriads of distances when facility locations are restricted to a finite set of candidates. An extensive bibliography of location theory literature is given by Francis and Goldstein [2] and a literature review of location in continuous space by Wesolowsky [9]. Convexity properties of the distance functions may be important to the user of continuous space location models. It is in response to this need that we have included a discussion of convexity properties.

2. Design of the Study

Ohio city distances in miles were derived from data supplied by the Ohio Department of Transportation. The data consisted of shortest path distances between each of a selected number of centroids in a given city. The rectangular coordinates in miles of each centroid were also supplied. The selected centroids were chosen to give a "representative" spread of points within the particular urban boundaries. From these, fifteen points were chosen randomly for each city. These points were used to generate a sample of 105 actual road distances--one for the distance between each pair of points. Wisconsin road distances were calculated using published road maps. An orthogonal coordinate system was defined and the fifteen sample generating points were selected by randomly determining pairs of coordinates and discarding points which fell outside the study area. Distances and point coordinates were measured directly from the maps in eighths of inches. This introduces a difference in the scale of measurement between the Ohio and Wisconsin data. This scale difference explains the larger magnitude of values of the goodness-of-fit criteria AD_f and SD_f in Tables 1-5 for Wisconsin data relative to Ohio data.

Nordbeck [7] states that a division of a city is necessary when modeling road distances if a river effects a separation of the road net. This is true since the road net between a pair of points, one on each side of the river, is not homogeneous when travel must be routed via bridges. To explore this concern in the study, a random sample of fifteen of the centroids available for Toledo were chosen from those to the west of the Maumee River which divides the city. This constitutes the seventh set of sample data which will be designated the Toledo Subarea.

Fitting Criteria

As in the previous study, two criteria of goodness-of-fit were used to measure the accuracy of the functions. The first is the minimization of a sum of absolute deviations given by

$$AD_f = \sum_{j<t} |d_f(a_j, a_t) - A(a_j, a_t)|,$$

where $j=1,\ldots,14$; $t=j+1,\ldots,15$ and $A(a_j, a_t)$ is the actual road distance between points a_j and a_t. This criterion requires estimation of greater actual distances relatively more accurately than shorter distances. The second criterion is the minimization of a sum of squares given by

$$SD_f = \sum_{j<t} \{[d_f(a_j, a_t) - A(a_j, a_t)]/\sqrt{A(a_j, a_t)}\}^2.$$

Division by $\sqrt{A(a_j, a_t)}$ normalizes the squared deviation and renders this criterion more sensitive than the first to large values of $|d_f(a_j,a_t) - A(a_j,a_t)|$ in relation to $A(a_j, a_t)$. In this way SD_f offers an alternative notion of goodness-of-fit.

Parameter Definition

The parameters of the functions were chosen as those which best fit the given criterion for the given sample. The "best" parameter values for each sample depend on the criterion used. Computer programs were developed to perform exhaustive searches for optimal parameters within chosen intervals.

For $d(q, r; k, p, p)$ the intervals of search were $k\epsilon[0.80, 2.29]$ and $p\epsilon[0.90, 2.29]$. This choice was made to allow for the best-fit value of each parameter to fall outside the expected range from 1 to 2. For d_1 the intervals of search were $k\epsilon[0.8, 1.25]$ and $\theta\epsilon[0^o, 90^o]$, while for d_2 the search interval was $k\epsilon[0.8, 2.29]$. The parameter search in the 3-parameter distance models was conducted in the following way, for each of the 2 criteria.

First the criterion function was evaluated at every grid-point in a unit cube in the parameter space, using a grid-width of 0.1. In almost all cases the minimum value of the criterion function was obtained at an interior point of the cube. Where this did not happen, the process was repeated after shifting the cube in the appropriate direction, until an interior minimum on the 0.1-grid was formed.

At this point a search was conducted on the grid-points in a cube with side-length 0.2 centered on the best point in the coarse grid, using a grid-width of 0.01. When the minimum value of the criterion function was obtained at an interior point of the cube, the parameter search was terminated. This happened with both criteria for all geographical areas, when d_5 was employed. With d_4, however, in many cases the search over the fine grid had to be repeated several times. Each time the cube was shifted in an appropriate direction until an interior minimum was found.

The search procedure employed does not ensure that a global optimum (to 2 decimal points) was found. However, several of the parameter values were verified by doing a total enumeration on a relatively coarse grid (.1). We feel confident, therefore, that the reported parameter values in most cases are optimal. Where this is not the case, the reported figures are probably near-optimal, at least with respect to the value of the criterion function.

3. Statistical Tests

The differing estimating accuracy of the distance functions can be associated with statistical significance using the t-test for matched pairs. Define $X_f(a_j, a_t) = |d_f(a_j, a_t) - A(a_j, a_t)|/\sqrt{A(a_j, a_t)}$ and the difference $dX_{gf}(a_j, a_t) = X_g(a_j, a_t) - X_f(a_j, a_t)$. Each such difference is between two observations on the same selected road distance. Division by $\sqrt{A(a_j, a_t)}$ is used to accomplish homoscedasticity. This was necessitated since the variation in $|d_f(a_j, a_t) - A(a_j, a_t)|$ tends to be directly related to the magnitude of $A(a_j, a_t)$. These definitions lead naturally to a test of differences between means based on matched pairs. This test, as opposed to others that might have been used, is unaffected by the lack of independence of the error in the distance functions. The test statistic is $t_{gf} = \overline{dX}_{gf}/S_{\overline{dX}_{gf}}$, where

$\overline{dX}_{gf} = \sum_{j<t} dX_{gf}(a_j, a_t)/n$, $S_{\overline{dX}_{gf}}$ is the standard error of the difference and $n=105$. The necessary assumption that the population of differences is normally distributed in order that t_{gf} has the t-distribution with n-1 degrees of freedom appears reasonable in this context of estimation errors.

Consider the first criterion of goodness-of-fit. If d_g actually produces values of AD_g lower than that for AD_f then t_{gf}, which is made up of terms of AD_g and AD_f, has a negative expectation. Said another way, if d_g is "more accurate" than d_f then the average value of $X_g(q,r)$ should be less than that for $X_f(q,r)$. As the computed value of $t_{gf} \to -\infty$, the weight of sample evidence favors rejection of the null hypothesis $E[X_g(q,r)]-E[X_f(q,r)] = 0$ and lends statistical significance to the negative value of $\overline{dX}_{gf}$. To give substance to this notion a significance probability is reported as the area under the t-distribution with 104 degrees of freedom in the interval $(-\infty, t_{gf})$. Although a t-test could be devised analogously to be associated with the second criterion of goodness-of-fit this was not done in order to avoid undue proliferation of statistical tests.

4. Presentation and Analysis of Results

The results of fitting the five distance estimating functions to the seven sample sets of urban and rural road distances are summarized in Tables 1-5. The statistical values associated with the fitting criteria are given in Table 6. Certain observations can be made. The goodness-of-fit of d_4 to the actual road distances was superior to that of d_5, the other three parameter function, in every case except for the rural Wisconsin data. Statistical significance accompanied this superiority in the Columbus, Milwaukee, Toledo and Toledo Subarea samples. d_4 includes d_1, d_2 and d_3 as special cases and therefore estimates distance at least as accurately. This structural superiority led to statistical significance at every opportunity except for d_2 and d_3 in the Madison and rural Wisconsin samples. The fitted parameters of d_4 do not satisfy convexity conditions for any but the Madison and Milwaukee data sets. This result may frustrate the use of d_4 in facilities location objective functions which are part of urban location models. A practical remedy would be to fit the parameters of d_4 under the condition $p \geq s$. The simplifying assumption that $p = s$ suggested by previous results [6] does not seem justified for the Ohio data. The fitted parameter values for d_3 and d_5 satisfy the respective convexity conditions in every case.

The rectangular distance function d_1 does not fare well in relation to the alternatives. In most cases this inferiority was substantiated by clear statistical significance. Indeed, only for the Milwaukee sample did the accuracy of d_1 surpass even that for the Euclidean function d_2. The computationally convenient rectangular distances may therefore provide a relatively poor fit to a variety of urban data bases.

Table 1

Minimizing Parameters of d_1*

Sample	AD_1	k	θ	SD_1	k	θ
Canton	123.05 (136.71)	1.04(0.95)	21	14.11 (17.86)	1.03(0.96)	20
Columbus	86.07 (88.35)	1.02(0.99)	5	10.53 (10.91)	1.01(1.00)	5
Madison	1015.96(1057.83)	0.98(0.97)	4	157.49(163.42)	0.98(0.98)	3
Milwaukee	628.95 (727.10)	0.95(0.95)	5	66.86 (87.83)	0.94(0.94)	6
Rural Wis.	508.42 (515.81)	1.05(1.05)	3	94.95 (96.38)	1.04(1.[illegible])	3
Toledo	99.73 (114.83)	0.97(0.99)	15	14.47 (18.30)	0.98(0.[illegible])	12
Toledo Sub.	69.60 (76.12)	0.97(0.94)	15	11.42 (13.98)	0.97(0.94)	20

*Parenthetical values correspond to the original axes.

Table 2

Minimizing Parameters of d_2

Sample	AD_2	k	SD_2	k
Canton	78.39	1.20	7.42	1.22
Columbus	80.20	1.27	8.78	1.28
Madison	869.40	1.25	125.73	1.25
Milwaukee	714.83	1.16	72.17	1.18
Rural Wis.	480.19	1.35	72.90	1.34
Toledo	74.13	1.21	8.70	1.24
Toledo Sub.	49.17	1.21	7.18	1.23

Table 3

Minimizing Parameters of d_3*

Sample	AD_3	k,p	SD_3	k,p
Canton	65.62*	1.16,1.49	6.03*	1.18,1.56
Columbus	68.54	1.18,1.47	6.86	1.18,1.45
Madison	830.34	1.16,1.48	113.24	1.18,1.56
Milwaukee	508.81*	1.03,1.30	46.31*	1.07,1.35
Rural Wis.	452.74*	1.24,1.45	70.45	1.29,1.65
Toledo	71.69	1.18,1.74	8.41*	1.20,1.73
Toledo Sub.	48.13*	1.18,1.78	7.07*	1.21,1.81

*Entries correspond to original axes or axes rotated through the angle given in Table 1, whichever case yields the lower criterion value. An asterisk denotes that the result is for rotated axes.

Table 4

Minimizing Parameters of d_4*

Sample	AD_4	k,p,s	SD_4	k,p,s
Canton	60.34*	1.43,1.66,1.78	4.10	1.53,1.81,1.99
Columbus	53.76	1.45,1.38,1.52	4.39	1.49,1.41,1.57
Madison	824.20	1.01,1.56,1.51	113.20	1.14,1.54,1.53
Milwaukee	508.81*	1.06,1.30,1.30	46.29*	1.03,1.34,1.33
Rural Wis.	447.10	1.56,1.50,1.59	70.24	1.37,1.65,1.68
Toledo	54.90	1.56,1.68,1.90	4.95	1.56,1.70,1.92
Toledo Sub.	34.08*	1.55,1.88,2.12	2.69	1.64,2.35,2.70

*See note for Table 3.

Table 5

Minimizing Parameters of d_5

Sample	AD_5	m_1 , m_2 , m_3	SD_5	m_1 , m_2 , m_3
Canton	63.37	1.47,-0.41,2.02	5.17	1.47,-0.29,1.92
Columbus	70.99	1.46,-0.99,1.78	7.52	1.50,-0.04,1.81
Madison	855.30	1.65,-0.01,1.46	120.00	1.64, 0.00,1.44
Milwaukee	657.70	1.64,-0.09,1.27	62.05	1.56,-0.11,1.31
Rural Wis.	430.60	1.82,-0.48,1.90	66.80	1.81,-0.36,1.85
Toledo	67.90	1.52,-0.23,1.37	8.01	1.57,-0.21,1.47
Toledo Sub.	47.54	1.58,-0.09,1.43	6.77	1.70,-0.03,1.44

Table 6

Values of $\overline{dX}_{gf}$ and Significance

Probabilities (Parenthetically)

Canton Sample

f \ g	2	3	4	5
1	-.09 (.00)	-.13 (.00)	-.15 (.00)	-.14 (.00)
2		-.03 (.00)	-.06 (.00)	-.05 (.00)
3			-.03 (.00)	-.01 (.18)
4				.02 (.87)

Columbus Sample

f \ g	2	3	4	5
1	-.01 (.38)	-.04 (.01)	-.09 (.00)	-.03 (.12)
2		-.03 (.00)	-.08 (.00)	-.02 (.02)
3			-.05 (.00)	.01 (.75)
4				.06 (1.00)

Table 6 - Continued

f		g: 2	3	4	5
	1	-.20 (.00)	-.23 (.00)	-.23 (.00)	-.22 (.00)
	2		-.03 (.25)	-.03 (.20)	-.01 (.29)
f	3	Madison Sample		-.00 (.45)	.01 (.62)
	4				.02 (.65)

f		g: 2	3	4	5
	1	.06 (.81)	-.05 (.11)	-.12 (.00)	.01 (.60)
	2		-.11 (.00)	-.18 (.00)	-.05 (.12)
f	3	Milwaukee Sample		-.07 (.00)	.07 (.95)
	4				.14 (1.00)

f		g: 2	3	4	5
	1	-.04 (.22)	-.07 (.01)	-.07 (.03)	-.10 (.05)
	2		-.03 (.09)	-.03 (.09)	-.06 (.02)
f	3	Rural Wisconsin Sample		.00 (.53)	-.03 (.26)
	4				-.03 (.24)

Table 6 - Continued

Toledo Sample

f \ g	2	3	4	5
1	-.07 (.00)	-.07 (.00)	-.12 (.00)	-.08 (.00)
2		-.00 (.15)	-.06 (.00)	-.01 (.06)
3			-.06 (.00)	-.01 (.17)
4				.05 (.99)

Toledo Subarea Sample

f \ g	2	3	4	5
1	-.05 (.00)	-.06 (.00)	-.15 (.00)	-.06 (.00)
2		-.00 (.20)	-.10 (.00)	-.01 (.01)
3			-.09 (.00)	-.01 (.11)
4				.09 (1.00)

The Euclidean function which models road distances as having no directional bias emerges as preferable to d_1 on the basis of accuracy and convenience. Only the inflation factor k must be estimated and an optimal coordinate system need not be identified. If this optimality were not incorporated into d_1 through the angle θ and the original axes were used, the dominance of d_2 over d_1 would be total. The values of k were generally greater for these sampled distances than for the inter-city distances of the previous study.

The two-parameter function d_3 is not decidedly inferior to either d_4 or d_5, each of which requires the estimation of three parameters. The values of p for d_3 even for the cases of rotated axes do not appear to be consistently close to unity. This would further substantiate the lack of a strong rectangular bias in the road networks of the chosen study areas. Addition of the parameter θ, as in d_1, to provide an optimal coordinate system could only improve the accuracy of d_3. This was not done here in order to keep the computational burden within tolerable limits.

The "quadratic form" function d_5 proves to be relatively accurate. The absolute value of m_2 is typically close to zero. This implies that the major and minor axes of the ellipse defined by $d_5(q,0) = 1$ are generally close to coinciding with the coordinate axes used in the study. Since $m_1 \neq m_3$, the locus of points satisfying $d_5(q,0) = 1$ is indeed an ellipse. The parameters for d_5 using rural distances provide the greatest deviation from these generalizations. For this sample, $m_1 \approx m_3$. Since $m_2 = -0.48$ and -0.36, the major axis of the ellipse points in a southwest-northeast direction. Additionally, d_5 best estimates the rural distances under both criteria. The directional bias afforded by d_5 may have special application to such rural settings.

The rural distances sample leads to the greatest value of k in each of d_1, d_2 and d_3. This sample does not otherwise lead to distinguishably different parameter values for d_1 through d_4 in relation to the other samples. The values of p show no tendency toward 1 or 2 in d_3, while the values of p and s which are similar for d_4 also fall almost midway between 1 and 2. The Toledo sub-area sample allowed for lower values of each AD_f and SD_f than those for the Toledo sample which includes some points on opposite sides of the Maumee River. The parameter values of the functions are different for the two Toledo samples.

In general, the parameter values for each function were different for different samples and for each criterion of goodness-of-fit. In the previous study, parameter values were relatively the same for each criterion. This signals the importance of fitting an empirical distance function to the particular area of study when a premium is placed on accuracy. There is a robustness in the relative accuracy of the functions but the analyst must tailor the parameter values to the study area and according to a useful criterion of goodness-of-fit. In order to illustrate how sensitive the criterion functions are to changes in the parameter values, we give the following two tables of criterion values in a coarse grid centered at the optimal point. In both cases the geographical area is Canton, the criterion is SD_f, and the grid-width is 0.1.

Comparing the two tables, it appears that for d_5 the second criterion is convex in the parameters m_1, m_2, and m_3, and is only slightly sensitive to changes in them. For d_4 and fixed k the second criterion seems to give rise to a surface like a flat-bottomed valley with very steep sides. This difference accounts for the relative difficulty in estimating the parameters of the second model. For this model, Table 7 also indicates that parameter values in an extended region may give a near-optimal fit, as long as the proper balance is maintained between p and s.

Table 7

Values of SD_4 About k=1.6,
p=1.7, s=1.9 where SD_4 = 4.8

k=1.5

p \ s	1.8	1.9	2.0
1.6	8	46	106
1.7	18	8	43
1.8	110	17	7

k=1.6

p \ s	1.8	1.9	2.0
1.6	4.9	23.	71.
1.7	47.	4.8	20.
1.8	187.	45.	4.8

k=1.7

p \ s	1.8	1.9	2.0
1.6	15	10	43
1.7	94	15	8
1.8	288	90	15

Table 8

Values of SD_5 About the Optimum
m_1=1.5, m_2=-0.3, m_3=1.9 where SD_5 = 5.2

m_1=1.4

m_2 \ m_3	1.8	1.9	2.0
-0.4	7.5	6.7	6.2
-0.3	6.5	5.9	5.7
-0.2	5.8	5.5	5.5

m_1=1.5

m_2 \ m_3	1.8	1.9	2.0
-0.4	5.7	5.3	5.3
-0.3	5.4	5.2	5.5
-0.2	5.5	5.6	6.0

m_1=1.6

m_2 \ m_3	1.8	1.9	2.0
-0.4	6.3	6.4	6.7
-0.3	6.7	7.0	7.5
-0.2	7.4	7.9	8.9

5. Conclusions

Empirical distance functions are suggested for improved accuracy in predicting actual road distances between two points in an area under study. Parameter values may be estimated upon defining a meaningful goodness-of-fit criterion. The standard assumption that urban distances are rectangular is not supported by the results reported here. The Euclidean distance is more convenient to use and appears to estimate urban distances more accurately unless the road network is endowed with a strong rectangular bias. The fifth distance model produced reasonably accurate results. However this function, with its unique directional bias, would seem to be especially suitable where the road network is not highly developed--such as in certain low population density regions. In addition, the parameters of d_5, which always satisfied convexity conditions, were more easily estimated than those for d_4, the other three-parameter function. d_4 proved to be most accurate for the urban data sets. However, the empirically determined parameter values typically failed to satisfy convexity conditions. A remedy would be to enforce the convexity conditions in the parameter estimation. d_3 gave relatively accurate results although it was shown to be inferior to the more general d_4 through statistical analysis. If a three parameter function is to be considered, introducing the parameter θ into the definition of d_3 to optimally orient the coordinate system would combine convexity and ease of parameter estimation with accuracy.[1]

[1]Robert Love's work was supported by U.S. Army Contract No. DAAG29-75-C-0024 and a grant from the Graduate School of the University of Wisconsin-Madison. James Morris' work was supported by a grant from the Graduate School of Kent State University. The assistance of the Ohio Department of Transportation, especially Charles Gebhardt and Charles Groves, in supplying the Ohio data for this study is gratefully acknowledged. Henrik Juel provided helpful assistance in completion of the study. Dean Wichern provided helpful discussions of the statistical analysis. A special note of thanks is extended to the anonymous referee whose criticisms of an earlier draft led to substantive improvements in the paper.

APPENDIX

Convexity Properties

Proof of the convexity of $d_4(q,r)$ when $p \geq 1$, $p \geq s$, and $k \geq 0$:

It is sufficient to prove that $\left[\sum_{i=1}^{2} |q_i - r_i|^p\right]^{1/s}$ is convex. It is known that $d(q, r; 1, p, p)$ is convex. Let $g(y) = y^b$, $y \geq 0$ and $b \geq 1$. Then g is an increasing convex function of y. Letting $b=p/s$ we can write $\left[\sum_{i=1}^{2} |q_i - r_i|^p\right]^{1/s} = g(d(q, r; 1, p, p))$ which is convex for $p \geq 1$, $p \geq s$, since an increasing convex function of a convex function is convex.

This extends and generalizes an earlier result by Cooper [1] that $\phi(x,y) = \beta_j[(x_j - x)^2 + (y_j - y)^2]^{K/2}$ is a convex function of (x,y) for $k > 1$ and $\beta_j \geq 0$ since the above proof neither requires that x_j and y_j be constants nor that $p=2$.

Proof of the convexity of $d_5(q,r)$ when M is positive definite:

Let $(q_1, r_1, q_2, r_2) = (x_1, x_2, x_3, x_4) \equiv x'$ and $y' \equiv (y_1, y_2, y_3, y_4)$. Then

$$d_5(q,r) = d_5(x) = [m_1(x_1 - x_2)^2 + m_2(x_1 - x_2)(x_3 - x_4) + m_3(x_3 - x_4)^2]^{\frac{1}{2}} .$$

To prove the convexity of d_5 we must prove that for any x and y and any non-negative numbers α, β, where $\alpha + \beta = 1$, the inequality

$$d_5(\alpha x + \beta y) \leq \alpha d_5(x) + \beta d_5(y)$$

must hold.

Since M is positive definite,

$$0 \leq [d_5(\alpha x + \beta y)]^2 = \left[\alpha \begin{bmatrix} x_1 - x_2 \\ x_3 - x_4 \end{bmatrix} + \beta \begin{bmatrix} y_1 - y_2 \\ y_3 - y_4 \end{bmatrix}\right]' M \left[\alpha \begin{bmatrix} x_1 - x_2 \\ x_3 - x_4 \end{bmatrix} + \beta \begin{bmatrix} y_1 - y_2 \\ y_3 - y_4 \end{bmatrix}\right]$$

$$\leq \alpha \begin{bmatrix} x_1 - x_2 \\ x_3 - x_4 \end{bmatrix}' M \begin{bmatrix} x_1 - x_2 \\ x_3 - x_4 \end{bmatrix} + \beta \begin{bmatrix} y_1 - y_2 \\ y_3 - y_4 \end{bmatrix}' M \begin{bmatrix} y_1 - y_2 \\ y_3 - y_4 \end{bmatrix} .$$

See Hadley [4, p. 85].

Taking the square root of both sides of the second inequality, and noting that $(a_1 + a_2)^{\frac{1}{2}} \leq a_1^{\frac{1}{2}} + a_2^{\frac{1}{2}}$ when $a_1, a_2 \geq 0$, we can write

$$d_5(\alpha x + \beta y) \leq \left[\alpha \begin{bmatrix} x_1 - x_2 \\ x_3 - x_4 \end{bmatrix}' M \begin{bmatrix} x_1 - x_2 \\ x_3 - x_4 \end{bmatrix}\right]^{\frac{1}{2}} + \left[\beta \begin{bmatrix} y_1 - y_2 \\ y_3 - y_4 \end{bmatrix}' M \begin{bmatrix} y_1 - y_2 \\ y_3 - y_4 \end{bmatrix}\right]^{\frac{1}{2}}$$

$$\leq \alpha d_5(x) + \beta d_5(y),$$

which is the desired result.

REFERENCES

1. COOPER, L., "An Extension of the Generalized Weber Problem," Journal of Regional Science, Vol. 8, No. 2 (1968), p. 181.

2. FRANCIS, R. L. and GOLDSTEIN, J. M., "Location Theory: A Selective Bibliography," Operations Research, Vol. 22, No. 2 (1974), p. 400.

3. GINSBURGH, V. and HANSEN, P., "Procedures for the Reduction of Errors in Road Network Data," Operational Research Quarterly, Vol. 25, No. 2 (1974), p. 321.

4. HADLEY, G., Nonlinear and Dynamic Programming, Addison-Wesley, Reading, Mass., 1964.

5. KOLESAR, P., WALKER, W. and HAUSNER, J., "Determining the Relation Between Fire Engine Travel Times and Travel Distances in New York City," Operations Research, Vol. 23, No. 4 (1975), p. 614.

6. LOVE, R. F. and MORRIS, J. G., "Modelling Inter-City Road Distances by Mathematical Functions," Operational Research Quarterly, Vol. 23, No. 1 (1972), p. 61.

7. __________ and ____________, "Solving Constrained Multi-Facility Location and Problems Involving ℓ_p Distances Using Convex Programming," Operations Research, Vol. 23, No. 3 (1975), p. 581.

8. NORDBECK, S., "Computing Distances in Road Nets," Regional Science Association: Papers XII, European Congress (1963), p. 207.

9. WESOLOWSKY, G. O., "Location in Continuous Space," Geographical Analysis, Vol. 5, No. 2 (1972), p. 95.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (6) 1719	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER (9) Technical
4. TITLE (and Subtitle) MATHEMATICAL MODELS OF ROAD TRAVEL DISTANCES.		5. TYPE OF REPORT & PERIOD COVERED Summary Report, - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Robert F./Love and James G./Morris		8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 5 (Mathematical Programming and Operations Research)
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE (11) February 1977
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 25p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(14) MRC-TSR-1719

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Estimates	Point	Rural
Distances	Coordinates	Statistical
Road	Functions	Parameter
Network	Urban	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Management science models often require estimates of distances between points in a road network based on the point coordinates. These estimates are typically derived using rectangular or Euclidean distance functions. The present investigation employs more general functions to estimate samples of urban and rural road distances. Proofs of conditions for the convexity of the functions are given. Statistical comparisons are made of prediction accuracy in order to establish relative merit. Sensitivity of the best parameters of the functions with respect to different geographical areas suggests the importance of using empirical distance functions.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

221 200

UNCLASSIFIED LB

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)